DOCUMENT RESUME

ED 171 794 TM 009 438

AUTHOR MacKay, A.
TITLE Evaluation of an Elementary School French Language Program: Impact Over a Six Year Period.
PUB DATE Apr 79
NOTE 20p.; Paper presented at the Annual Meeting of the American Educational Research Association (63rd, San Francisco, California, April 8-12, 1979)

EDRS PRICE MF01/PC01 Plus Postage.
DESCRIPTORS *Audiolingual Skills; Curriculum Development; Curriculum Evaluation; Educational Assessment; Foreign Countries; Formative Evaluation; *French; Intermediate Grades; *Language Programs; *Program Evaluation; *Second Language Learning
IDENTIFIERS Canada

ABSTRACT

Four evaluation studies were conducted of French as a second language program in grades 4, 5, and 6 in a large urban school district in western Canada. As a result of the 1972 study, which revealed major discrepancies between intended and observed outcomes, especially in oral skills, a major objectives-based curriculum was initiated. As part of the implementation of the revised curriculum: (1) objectives for each grade level stated in measurable terms; (2) time allocated for instruction increased from 90 to 150 minutes each week; (3) inservice teacher education was provided; and (4) articulation with the French program in grades 7 through 12 was coordinated. Three program evaluations, based on the new curriculum, were undertaken. In general, program evaluation showed improved performance in oral skills. Audio comprehension also improved in grades 4 and 5; but, in grade 6, audio comprehension tests favored the old program pupils. Side effects studies were not conducted. Inferences drawn from the study included the need for coordination between evaluators, program developers, and policy makers; the importance of the timing of evaluation studies; the high cost and difficulty of undertaking process evaluation; and the need for more standardized French achievement tests for Canadian elementary school students. (MH)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED171794

U S DEPARTMENT OF HEALTH,
EDUCATION & WELFARE
NATIONAL INSTITUTE OF
EDUCATION

THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL NATIONAL INSTITUTE OF EDUCATION POSITION OR POLICY

EVALUATION OF AN ELEMENTARY SCHOOL FRENCH LANGUAGE PROGRAM: IMPACT OVER A SIX YEAR PERIOD

A. MacKay
Centre for Research in Teaching
Faculty of Education
University of Alberta
Edmonton, Alberta, Canada

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY

A. MacKay

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC) AND USERS OF THE ERIC SYSTEM"

IM009 438

Paper presented at the annual meeting of the American Educational Research Association, San Francisco, California, April, 1979.

Introduction

After a dozen years of extensive activity by program evaluators a period of reflection and stock-taking has begun. As meta-evaluation replaces mere run of the mill evaluation, and as a large number of scholar — practitioners engage in consideration of appropriate standards for evaluation, the stage is set for retrospective, critical analyses of the impact of evaluation studies, their connection with policy-making, and their relation to curriculum development projects.

The series of studies reported in this paper and a short history of some of the events surrounding the preparation and delivery of these particular evaluation reports provide some data which will be treated as if it had been collected in an exploratory study of the following general questions:

1. What impact can evaluation studies have upon curriculum development and implementation?
2. What are some of the problems associated with implementation of a new second-language curriculum in an environment which may not be fully supportive of such a curriculum?

The series of evaluations took place in a large urban school district in one of the western provinces of Canada. The district had operated a program of French as a second language in Grades 4, 5, and 6 since 1960. In 1972, an evaluation of pupil performance, especially in oral production, revealed major discrepancies between intended and observed outcomes. As a result of that evaluation report, a major curriculum development project was launched with local, provincial and federal support. This project produced an objectives-

based curriculum which was implemented in Grades 4, 5, and 6 in the years 1975 to 1978.

At the end of the school years 1976, 1977, and 1978 the program was evaluated in grades 4, 5, and 6 respectively.

This paper presents the following information: (1) a description of the setting in which the studies took place; (2) a description of each of the four studies and a summary of the conclusions and recommendations; (3) a summary of the other relevant events which surrounded the evaluation studies (This is incorporated as "Next Steps" in conjunction with the summaries of the four reports.); (4) a tentative assessment of the impact of the studies; and, (5) an identification of some of the problems associated with this kind of curriculum project.

The Canadian Scene

Under the terms of the British North America Act (1867) and of other acts which figured in the establishment of provinces entering Canadian Confederation after 1867, a clear assignment of control over education was made to the provincial legislatures. However, during its first century a number of exceptions to the general constitutional rule have become part of the educational scene in Canada. The education of native Indians and Eskimos, of children of members of the armed forces, and, of students enroled in certain types of vocational and technical programs has had a federal "presence" through departments of Indian Affairs, National Defence, and Labour. During the past two decades a number of agreements between the "feds" and the provinces have been negotiated with respect to post-secondary education, and

of direct relevance to this paper, second language and bilingual instruction. A group called the Council of Ministers of Education, comprised as the name suggests of the ministers of education from each of the ten provinces, has acted as the negotiating body for the provinces. For language instruction, the office of the Secretary of State (a ministry of the Government of Canada) has represented the federal presence. The negotiations have led to a series of agreements about joint funding of various programs which have had a second-language or a bilingual flavor. It should be noted that, not unexpectedly, when one speaks of bilingualism in Canada at the level of provincial-federal relationships one is, of course, referring to English and French. While some provincial governments support instruction in languages other then English or French, the inter-provincial (a peculiarly Canadian term?) program activities have been focussed on French and English. By the way, the agreement under which the program described in this paper received some federal input was to expire at the end of 1978-79 and there are no clear indications about the future status of such arrangements.

Against this constitutional background, the history of the elementary French project is briefly as follows. From 1960 to 1972 approximately twenty minutes per day were allocated to instruction in Grades 4, 5, and 6 French, using a curriculum (Parlons Français) which was partly based on televised instruction. The program was entirely supported by the school district with the same level of provincial support it received for the various other programs offered in its schools. The 1972 evaluation led to a curriculum development

project supported by the federal and provincial governments as a special project. The project produced a new program of French as a second language, initially aimed at Grades 4, 5, and 6; but also including provisions for a new program covering grades 7 to 12. The new program was implemented on a phased basis during the years 1975 to 1978 in grades 4 to 6 and an evaluation was conducted, as mentioned above, of grade 4 (1976), grade 5 (1977), and grade 6 (1978). Costs associated with implementation of the new program during those years were shared, under a fairly complicated formula, by the federal and provincial ministries and the local school district. It is the future of the federal share (and any special provincial grants) which is uncertain as of the date of preparation of this paper (February, 1979).

The Evaluation Studies

The 1972 Study. The 1972 evaluation emerged from the concerns of central office supervisory staff as well as of personnel in the schools who expressed dissatisfaction with the Parlons Français program, particularly its heavy reliance on televised, pre-packaged lessons, and who were uncertain about the program's effectiveness particularly in terms of oral skills. The district had no research or evaluation office of its own and, therefore, approached the author of this paper with a request for an external evaluation. Shortly before this request was made, the district officials had already begun negotiations with the provincial department of education whose staff have had a fairly significant role, over many years, in evaluation of teachers, of schools, and of particular curriculum programs. Design of the

evaluation was assigned to the external evaluator and what emerged was a mixture of internal and external evaluation activities all of which were imbedded in a modified version of Stufflebeam's CIPP model. Contextual and product data collection was carried out by the external evaluator (then, as now, a staff member of the major provincial university), while input and process data, à la Stufflebeam's model, were obtained by school district personnel in cooperation with a department of education evaluation team. The decision to provide only loose coordination among the components implied by the CIPP model had major implications, as will be seen below, for the explanatory power of the evaluation findings.

In any case, data on audio comprehension were obtained for 1330 pupils in grades 4, 5, and 6. Oral production tests were administered on an individual basis by teachers who had been specially trained for this assignment. Other data sources included interviews of 72 teachers, and 18 administrators, as well as of either or both parents in 108 families. The classroom process data were collected by the department of education team using the same criteria and observational techniques that they were currently using in similar assignments across the province. Information on objectives and on the program's history were obtained by a document analysis and through interviews of key people in the school district.

A number of staff members from the district who were returning, in the late Spring of 1972, from sabbatical leaves, were also involved in the study as interviewers and oral language testers. General supervision of the project was undertaken by the external

evaluator, advised by a steering committee, and assisted by officials of the school district central office.

The main conclusions of the 1972 study were:

1. Teachers, administrators and parents were generally favorable towards the concept of French as a second language.
2. Pupils, including some who had moved into secondary grades by the time of the 1972 study, were favorable towards French as a second language.
3. Performance in audio comprehension was generally good.
4. Oral skill scores were generally "low"; i.e. below 50%.
5. The televised lessons were seen to be redundant and somewhat irrelevant by a majority of respondents.
6. There were some exceptions to the general pattern of poor performance in oral skill tests.

In connection with this last finding, it was impossible to link process data with product data in order to obtain tentative explanations as to why some classes seemed to enjoy superior learning environments. It was in this respect that the loose coordination of the department of education "process" team with the other components produced unhelpful results. We knew that some classrooms were doing well. What we did not know was what the teachers and pupils in these classrooms were doing that produced better results than we obtained in the majority of classes tested. The absence of a standard observation system and a rather heavy reliance on general impressions about classroom events made the process data useful mainly as corroborative rather than correlational or predictive information.

The recommendations from the 1972 study included the following:

1. The program objectives should be revised and clarified.
2. The school district should prepare a proposal for large-scale external support for development of its program.
3. Teachers should be assisted in acquiring skills needed for teaching French as a second language.

An excerpt from the final pages of the 1972 report seems to sum up the intentions and aspirations of the external evaluator.

> "This evaluation study probably constitutes the first phase of an over-all program for improvement of French language instruction. What it has best done is indicate some of the needs which ought to be met .... The next step open to the ... school system is to plan, implement, and evaluate a series of curriculum changes designed to provide a very favorable set of learning experiences in French for every elementary pupil."

Next Steps. The school district did respond to the recommendations and during the following two years began negotiations with provincial and federal government officials that led to preparation of a special project proposal. The special project was funded and during the school year 1974-75 a team of curriculum writers, under the direction of the supervisor of language instruction, produced the curriculum guide books and curriculum materials for a new program which had the following characteristics:

1. Objectives for each grade level stated in measurable terms.
2. Time allocated for instruction increased from an average of 90 minutes to an average of 150 minutes per week.
3. A program of in-service introduction of teachers to the new curriculum.

4. Articulation of the grades 4, 5, and 6 segment of the program with grades 7 to 12 segments.

The behavioral objectives pertained to audio comprehension, oral expression, and reading and written expression.

Towards the end of the 1974-75 school year, grade 4 classroom teachers were invited to adopt the new program on a voluntary basis. As a result of this invitation, 31 schools adopted the program in grade 4 while the remaining 32 elementary schools in the district stayed on the old program.

Part way through the 1975-76 year, the school district decided that an evaluation of the program's first year of operation should be conducted, and commissioned the author of this paper to design and carry out the evaluation.

The 1976 Study. The grade 4 study, unlike the 1972 study, was able to capitalize upon the fact that almost exactly equal numbers of grade 4 classes were using the old and new programs. The opportunity for a horse-race evaluation design was to be an important part of the background against which the 1976 study and those which followed it in 1977 and 1978 were carried out.

A random sample of eight schools from each of the two "naturally" occurring experimental conditions was selected and, within each grade 4 class, a random sample of 25% of the pupils were selected for testing in oral skills. A team of oral testers was comprised of university students who had worked in a nation-wide evaluation project. In addition, data on audio comprehension of pupils as well as the

opinions of teachers and administrators about implementation problems were collected.

The conclusions of the 1976 study were:

1. Teachers and administrators were generally positive in their opinions about the new program.
2. Implementation had created two major problems:
   a. reduction of time allocated to other subjects, and,
   b. difficulties related to scheduling a qualified French teacher in elementary schools which were, essentially, comprised of self-contained classrooms.
3. Pupils in the new program performed significantly better on the total test score (including audio comprehension) and on the subtests which focussed on oral skills.
4. An analysis of variance (nested design) showed a significant classroom effect even though old versus new program effects were significant.

The report went on to recommend that:

1. The new program be introduced in all grade 4 classes in the district.
2. Introduction of the grade 5 and 6 components of the new program should be carefully monitored.
3. A longitudinal study of side-effects of the new program should be designed.

Next Steps. Sometime before the 1976 evaluation was completed, the school board had adopted, upon recommendation of its French

language supervisor and one member of the senior administrative staff, a policy statement which made system-wide adoption of the new curriculum mandatory. In spite of some opposition from the superintendent of schools and some members of the district staff, the board unanimously adopted the policy which would, in their view, lead to high quality instruction in French as a second language.

Following receipt of the 1976 evaluation report, the board instructed its administrative staff to continue with implementation across all grade 4 classes and, also, to determine how the program could be continually monitored and whether a study of side-effects was appropriate. In fact, the administration did not give serious consideration to a side-effects study; but it did engage the same external evaluator to conduct a study of the grade 5 program.

The 1977 Study. The grade 5 evaluation design was simply a replication of that used, one year before, for the grade 4 program. The main conclusions were:

1. Teachers and administrators were positive about the new program.
2. The main implementation problems were associated with:
   a. reduction of time allocated to other subjects, and
   b. difficulties in scheduling resources (such as classrooms) and teachers.
3. Pupils on the new program performed significantly better than pupils in the old program in both audio comprehension and oral skills.

The report recommended that:

1. The new program should be introduced in all grade 5 classes in the district, and,
2. Side effects of the program should be examined.

Next Steps. Again, the school board instructed its staff to implement the program across all grade 5 classes and, also, to study the other recommendations contained in the report. To the knowledge of the external evaluator, no serious consideration was given to the question of side-effects; but the district did, again, engage the same researcher to conduct a grade 6 evaluation. Again, approximately 50% of the grade 6 classes had voluntarily adopted the new curriculum, just as their counterparts in grades 4 and 5 had done in the first two phases of implementation in 1975 and 1976.

The 1978 Study. The 1978 study was, to a large extent, a replication of the 1976 and 1977 studies -- with one major additional feature. During the 1977 study, a set of baseline data on pupils enrolled in the old program in grades 6, 9 and 12 was obtained. The decision to collect such data was premised on the assumption that old program students (and classes) were becoming a vanishing breed and that unless some bench-mark data were obtained no useful standards for comparison would be available as the new program was phased into grades 6 through 12.

The conclusions were:

1. While schools labelled as new program grade 6 schools all

had 150 minutes of instruction or more per week, a number of so-called old program schools also had 150 allocated minutes per week and only one was as low as the 90 minutes which the old program would normally entail.

2. Teachers and administrators were favorable towards the program. Again, the twin problems of:
   a. time reduction in other subjects, and,
   b. scheduling of French classes,
   emerged.
3. There was a significant difference in oral skills, favoring new program schools.
4. Audio comprehension tests favored old program pupils.
5. Compared with the grade 6 baseline data (from 1977) pupils in both new and old program schools generally did better in all of the tests.
6. There was a significant classroom effect for all tests. (Note: In the grade 5 study in 1977, no classroom effects analysis was possible.)
7. A correlational analysis indicated internal consistency among test components.

The recommendations, repetitious and redundant though they may be, were:

1. The new program should be made available in all grade 6 classes which are able to provide necessary resources such as qualified teachers and appropriate classroom settings.
2. The side-effects of the program should be examined.
3. Given the classroom effects, the district should consider a study of classroom behavior of teachers and pupils.

Next Steps. As of the early part of the calendar year 1979, the new program has been implemented in almost all grade 6 classes. No action has been taken by the district with respect to the side-effects study or the classroom process study.

What Does It Mean?

The two general questions posed in the introductory section of this paper referred to: (1) impact of the evaluation studies, and (2) problems associated with this type of curriculum project. The data presented in the main body of the paper provide some case-study bases for attempting tentative, highly-limited answers to those two questions.

Impact. The relationship between evaluation and policy-making is obviously complex. Most recent writers reflecting upon their experience in program evaluation have felt the need to comment about the highly politicized nature of evaluation, about the failure of evaluation to have impact, or about the conversion of evaluative information to uses never bargained for by the evaluator. The data reported here are, perhaps, suggestive of several tentative conclusions about the interface between evaluators and policy-makers.

It is quite evident, in the series of events reported on above, that the policy-makers were never convinced of the need for either a side-effects study or a classroom process study. It should be noted that the evaluation reports did not provide strong evidence or arguments for either type of study. A critical reader of the reports might,

indeed, suggest that the recommendations about these two areas for research were added on, almost as afterthoughts, because of some personal or professional interests which influenced the external evaluator. Nevertheless, taken at face value, the absence of response by the policy-makers may suggest that their operating model of evaluation is Tylerian, i.e. goal-based, and limited in scope to the questions originally agreed on during negotiations about the evaluation design. Some might say that this response, if at all typical of policy-makers, is a problem-centered, "brush-fire" approach to system operation, and that, in its own right it may be a perfectly appropriate way of maintaining the system. In other words, for a large school district, the truly valuable and usable information may best deal with trouble spots or concerns that have surfaced in some significant ways and not with <u>potential</u> trouble areas that, for the moment, have no advocates, no injured parties, and, therefore, no clientele.

A second inference drawn from the information reported here is that the timing of evaluation studies is often critical. In this case, a major policy decision was made before the program, in a trial form, had been evaluated. Fortunately for the reconstructed logic of the policy decision, the evidence favored continuance and expansion of the program; but one wonders about what might have happened had the data on pupil performance been non-significant or "negative". Perhaps, and one is out on a long speculative limb here, the policy decision would have been reopened; perhaps the evaluation studies would have been discredited because of faulty design; perhaps the evaluation

report would not have been presented to the policy-makers; perhaps, perhaps, perhaps.

It should be mentioned, in connection with the question of timing, that in each succesive case, the external evaluation was not commissioned until, at the very earliest, the month of March in each of the school years involved. It seemed always to be assumed that an evaluation should be done, that the same external evaluator ought to be engaged to do it; but there was a lack of lead time which might have shaped the evaluation design in different ways or of involvement in curricular decisions which might have increased the impact of evaluations.

In retrospect, the 1972 evaluation report had major impact in that it accentuated the need for the school district to upgrade its French program. Out of it came the externally funded project which, as an event in this particular setting, led to some significant curriculum development and implementation activities. Perhaps the external funding, given its availability, would have been sought in any case; perhaps the evaluation report merely paralleled the school system's intentions. Cause and effect attributions are impossible in this case; but, at least there is correlational evidence in favor of an inference that some impact was felt.

In summary, the four evaluation studies seemed to have initial impact upon the series of curriculum development activities and continuing impact as support for what the district had decided it would do anyway. An alternative conclusion such as the district reversing or seriously modifying its policy if the evidence had been negative about the new program is like an untestable alternative hypothesis; we just don't know.

Curriculum development. Three main problem areas, which appeared from our vantage point to be relevant to scholar - practitioners in curriculum, were as follows:

1. The separation of evaluation from the on-going process of curriculum development.
2. The high cost and relative difficulty of studying process or program installation questions.
3. The lack of readily available tests of pupil achievement.

There is obviously little that is new in these conclusions. The prescriptive literature on evaluation has, for many years, urged evaluators and project developers to get together and to use each other's special knowledge and skill in all phases of program development. This advocacy literature implies that evaluation would be on the scene from the needs assessment stage, through program planning and trial runs, through revisions and reformulations, to examination of the finished product and of its effects on learners. In the case described here, program developers operated in the back rooms of the school system and no resources were ever provided for a possible trip back to the drawing board, which might have been called for by the evaluation studies. As a case in point, the story told here is one more example of the discrepancies between rational, planned program development and the bounded rationality of real life situations. Therefore, our first point is, merely, a reiteration of the point that numerous other researchers have made.

The relatively high cost of obtaining process data which in this instance would have required development of an appropriate

observation system, training of bilingual coders, and negotiations for access to classrooms, is not a small point. Although installation of the "independent variable" clearly ought to be examined as part of any major curriculum evaluation project, the resources in terms of skills and money needed for such work are usually not available. There is, one conjectures, too neat a separation between program evaluation and teacher evaluation. Some of us may have dwelt so often on the need to evaluate curricula independently of the teachers who implement them, or have played a political game of negotiating with teachers and administrators who are reluctant to have evaluators enter their classrooms, that we have unwittingly created a fiction. The fiction is that instruction and programs are separate domains and that curriculum can be evaluated without looking at live contexts, practising teachers, and learning pupils. Whatever the reasons may be, the fact was in the case reported above, that process evaluation was never seen as a feasible, necessary component of program evaluation. The proof of the pudding was in the eating not in the cooking.

The third problem area is, in Canada, an especially significant one. There are simply not enough standardized tests of French language performance for use in the elementary grades. The problem is accentuated (Excusez!) in a special program such as the one described here. The problem of obtaining suitable tests is one which haunts evaluators; but it should be equally important for the curriculum developers and teachers who work closely with the program. Some agency with skills in test development would fill an unmet need, at least on the Canadian scene, if they undertook a project in the

development of tests for use in French as a second language programs.

Conclusions

The very tentative conclusions and conjectures drawn from a case study of a particular series of evaluation studies, set in a particular context and subject area, are put foreward here as possible contributions to the area of meta-evaluation. The contributions are, possibly, in terms of one more example of the problems that confront evaluators, and of a source for questions for exploration under the one general question: What should have been done differently by the evaluator in the case reported above or in ones similar to it? On this note of enquiry, the story ends.

REFERENCES

MacKay, A. Evaluation of the French Language Program of Instruction in Grades 4 - 6 in a Metropolitan District, Summer, 1972.

MacKay, A. Evaluation of the Grade 4 French Language Program in Metropolitan Schools. Fall, 1976.

MacKay, A. Evaluation of the Grade 5 French Language Program in Metropolitan Schools, Summer, 1977.

MacKay, A. Evaluation of the Grade 6 French Language Program in Metropolitan Schools, Summer, 1978.

The titles of the reports have been changed to maintain the anonymity of the school district. Information on availability of the reports may be obtained by writing to the author of this paper.